 INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS

Help




&gt; Sample Design, Weights, Variance, &amp; Missing Data

Slide  
\$cpInfoCurrentSlide\$  
of \$cpInfoSlideCount\$

# Sample Design, Weights, Variance, and Missing Data

 [Download Transcript](#)  [Download Slides](#)  [Download Glossary](#)

 INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS


Help

> Introduction to Post Secondary Education > Module Objectives

00:01:12

## Module Objectives

- Describe the [weights](#) used in postsecondary education studies that must be applied to assure data are representative of the target population.
  - Summarize the postsecondary education studies sample designs as they relate to analysis weights that account for disproportionate probabilities of selection and estimation impact.
- Describe the major techniques for calculating estimates and their [standard errors](#) using NCES postsecondary sample survey data.
- Explain and illustrate how these techniques are applied in NCES postsecondary studies using selected software packages.
- Describe the missing data, or reserve codes, used in postsecondary education studies.

 INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS


Help

> Introduction to Post Secondary Education > Postsecondary Education Data Sampling Design

00:01:14

## Postsecondary Education Data Sampling Design

- [Two-stage design](#) (NPSAS)
  - Uses IPEDS as a sampling frame to sample postsecondary institutions
  - After obtaining enrollments lists from institutions, a sample of students is drawn
  - NPSAS is explicitly stratified by institution type and may include oversamples of some subgroups
  - Subsamples are selected for postsecondary longitudinal studies

 INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS


Help

> Introduction to Post Secondary Education > Postsecondary Education Data Sampling Design

00:00:22

## Postsecondary Education Data Sampling Design

- Institutions and students do not have equal probabilities for selection, violating assumptions related to [Simple Random Sampling](#) (SRS).
- Procedures are required to account for:
  - unequal probabilities for selection, and
  - incorrect variance approximation using statistical methods that assume that SRS was used.
- Weights are used to account for unequal probabilities of selection.

 INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS


Help

> Introduction to Post Secondary Education > Accounting for Disproportionate Probabilities for Selection: Weights

00:00:49

## Accounting for Disproportionate Probabilities for Selection: Weights

- NPSAS only includes one or two weights:
  - NPSAS:04 and later – one study member weight
  - NPSAS:00 and earlier – student interview weight and study member weight
- Longitudinal studies may have more than one weight depending on the number of follow-ups and/or data sources.

 INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS


Help

> Introduction to Post Secondary Education > Postsecondary Education Data Weights: An Example from B&B:93/03

00:00:56

## Postsecondary Education Data Weights: An Example from B&B:93/03

- Base year weight (WTA000)
  - Students interviewed at graduation (1992-93) in NPSAS:93. Includes approximately 11,200 graduates.
- Base year/3rd follow-up weight (WTB000)
  - Students interviewed at their tenth year after graduation (2003). Includes approximately 9,000 graduates.
- Panel weight (WTC000)
  - Students interviewed at graduation and first, third, and tenth year after graduation (1993, 1994, 1997, and 2003). Includes approximately 8,000 graduates.
- 1st follow-up/3rd follow-up Weight (WTD000)
  - Students interviewed at the first and tenth year after graduation (1994 and 2003). Includes approximately 10,400 graduates. (WTD000).

 INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS

Help

> Introduction to Post Secondary Education > Standard Error Calculation: Replication Techniques

00:01:03

**Standard Error Calculation in Postsecondary Education Data: Replication Techniques**

- This method calculates appropriate standard errors (SEs) using multiple weights and is the preferred method for calculating standard errors in postsecondary data.
- Select replicate weights that are associated with your analytic weight (e.g., WTA001 to WTA200 for weight WTA000).
- Most postsecondary studies replication weights use the Balanced Repeated Replicates (BRR) method.

ies

INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS

Help

> Introduction to Post Secondary Education > Weights on a Data File


00:00:42

## Weights on a Data File

### Balanced Repeated Replicate (BRR) Weights

**BPS:04/09**

BPS09PSU	Analysis replicate (PSU)
BPS09STR	Analysis stratum
WTC000	BPS:04/09 transcript weight
WTC001	BPS:04/09 bootstrap replicate weight - 1
WTC002	BPS:04/09 bootstrap replicate weight - 2
WTC003	BPS:04/09 bootstrap replicate weight - 3
WTC004	BPS:04/09 bootstrap replicate weight - 4
WTC005	BPS:04/09 bootstrap replicate weight - 5
↓	
WTC200	BPS:04/09 bootstrap replicate weight - 200

 INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS

Help

> Introduction to Post Secondary Education > Standard Error Calculation: Taylor Series Linearization

00:00:32

**Standard Error Calculation in Postsecondary Education Data: Taylor Series Linearization**

- This method uses primary sampling unit (PSU) and strata identifiers to calculate appropriate SEs.
- Select the identifiers that are associated with your main sampling weight (e.g., XXX\_PSU and XXX\_STR for weight WTA000 as the variable name usually includes 'PSU' or 'STR').
- Variables for Taylor Series Linearization are provided, but BRR weights are preferred.

ies

INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS

Help

> Introduction to Post Secondary Education > Weights on a Data File


00:00:43

## Weights on a Data File

### Taylor Series Weights

BPS:04/09

BPS09PSU	Analysis replicate (PSU)
BPS09STR	Analysis stratum
WTC000	BPS:04/09 transcript weight
WTC001	BPS:04/09 bootstrap replicate weight - 1
WTC002	BPS:04/09 bootstrap replicate weight - 2
WTC003	BPS:04/09 bootstrap replicate weight - 3
WTC004	BPS:04/09 bootstrap replicate weight - 4
WTC005	BPS:04/09 bootstrap replicate weight - 5
↓	
WTC200	BPS:04/09 bootstrap replicate weight - 200

 INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS

Help

> Introduction to Post Secondary Education > Missing Data and Imputation: Item Nonresponse and Missing Data

00:00:40

## Missing Data and Imputation: Item Nonresponse and Missing Data

- Why?
  - Respondent did not want to answer
  - Respondent broke off
  - Mortality
- Implication?
  - Those who responded to item are different in important ways from those who did not respond

ies

INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS


Help

> Introduction to Post Secondary Education > Missing Data and Imputation: Item Nonresponse and Missing Data (Continued)

00:01:14

**Missing Data and Imputation: Item Nonresponse and Missing Data (Continued)**

- A solution?
  - Imputation
    - Available information and some assumptions are used to derive substitute values for missing values in a data file
  - Generally, key items on most NCES files are imputed
- A warning!
  - When data are missing or not applicable, a [reserve code](#) will be placed in the field (e.g., -3 for a legitimate skip or -9 for missing). Leaving them there will throw off your estimates!

 INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS

Help


> Introduction to Post Secondary Education > Summary

00:00:25

## Module Summary

Now that you have completed this module, you should be able to:

- Describe the weights used in postsecondary education studies that must be applied to assure data are representative of the target population.
  - Summarize the postsecondary education studies sample designs as they relate to analysis weights that account for disproportionate probabilities for selection and estimation impact.
- Describe the major techniques for calculating estimates and their standard errors using NCES postsecondary sample survey data.
- Describe the missing data, or reserve codes, used in postsecondary education studies.

 INSTITUTE OF  
EDUCATION SCIENCES

NATIONAL CENTER FOR  
EDUCATION STATISTICS

Help

> Introduction to Post Secondary Education > Resources

00:00:18

## Module Resources

- [Weights](#)
- [Standard Errors](#)
- [Two-Stage Design](#)
- [Simple Random Sample](#)
- [Replication Techniques](#)
- [Taylor Series Linearization](#)
- [Reserve Codes](#)